

# Interactive Keyword-Based News Retrieval:

## Enhancing Content Discovery with NLP and Real-Time Data Processing

Ajay Raj Singh<sup>1</sup>, Venkata Akhil Mettu<sup>2</sup>, Dr. Gulhan Bize<sup>3</sup>  
Saint Peter's University, Data Science Institute, Jersey City, NJ, USA

### INTRODUCTION

We present an interactive news visualization system that transforms curated headline tags from the New York Times into dynamic word clouds. By integrating Google Trends data, our system enhances the contextual relevance of each keyword. This approach offers users an accessible way to explore evolving news topics with real-world importance.

### OBJECTIVE

This research aims to address the need for accessible news tools by developing an interactive word cloud dashboard that visualizes trending news topics. The primary objectives include:

- Creating a streamlined process for retrieving and analyzing top headlines from diverse sources
- Developing a visually engaging interface that presents complex news data in an accessible format
- Reducing information overload by emphasizing commonly used phrases and topics
- Improving news accessibility for underserved populations and those facing language barriers
- Enhancing user engagement with current events through interactive visualization

#### Target Audiences

Students	Researchers	Educators
Policy Makers	General Public	Media Analysts

### SYSTEM ARCHITECTURE

The Trending Topics Explorer implements a modular architecture that integrates data collection, processing, visualization, and user interaction components:



- Data Acquisition Layer:** Interfaces with news APIs to collect textual and video content
- Processing Layer:** Handles data cleaning, NLP tasks, and term extraction
- Storage Layer:** Maintains processed data for efficient retrieval
- Visualization Layer:** Generates the interactive word cloud using D3.js
- Presentation Layer:** Delivers the user interface for content exploration

News Assessment Scoring Formula: Final Score =  $\alpha \cdot \text{Base Score} + \beta \cdot \text{Trend Score}$

### RESULTS & IMPACT

#### News Sources

2+

Primary APIs integrated

#### Topic Coverage

100+

Daily trending topics

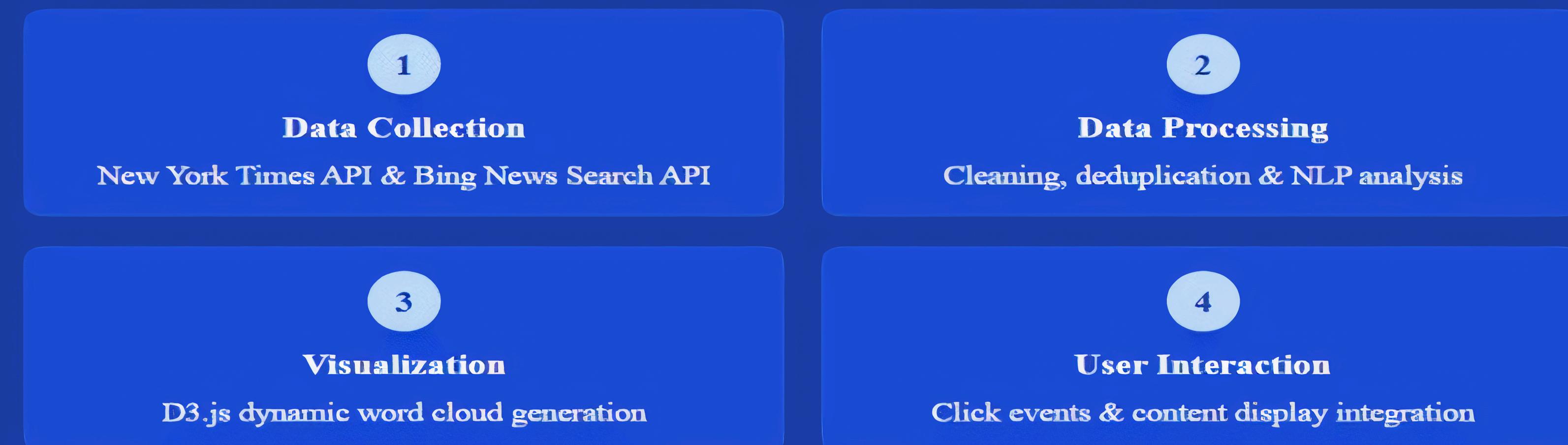
#### User Benefits

- Reduced information overload through visual summarization
- Increased engagement with current events
- More efficient discovery of relevant content
- Improved accessibility for diverse user groups

#### Educational Applications

- Fostering news literacy among students
- Providing a starting point for research projects
- Supporting current events discussions with visual aids
- Helping students track trends and develop critical analysis skills

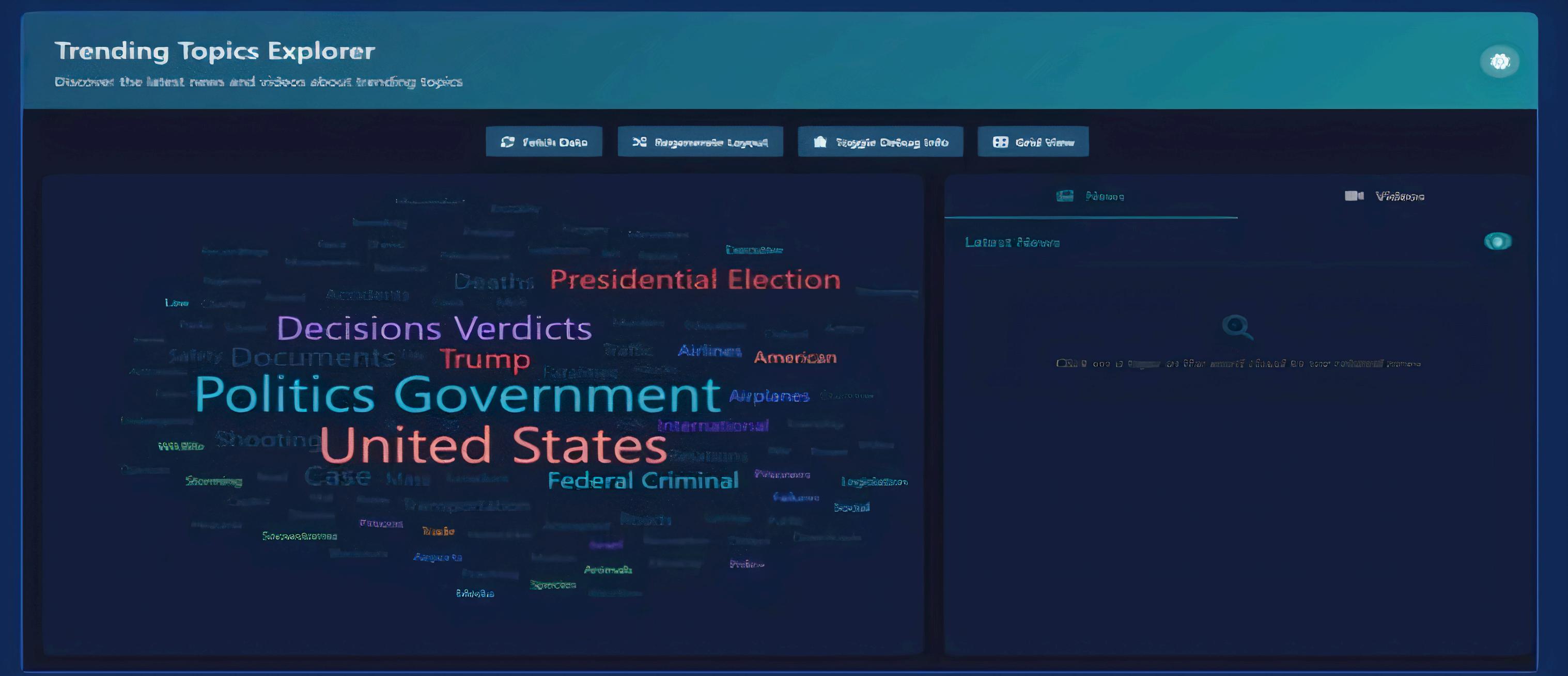
### METHODOLOGY & TECHNOLOGY STACK



The data collection process leverages multiple sources including the New York Times API for primary headlines and the Bing News Search API to enrich the dataset with additional articles and video content, providing diverse perspectives on trending topics.

Advanced NLP techniques are applied to process the collected data, including text normalization, stop word removal, term extraction, and frequency analysis to identify significant words and themes for visualization.

### INTERACTIVE VISUALIZATION



Try it yourself: [QR code](#)

### CONCLUSION & FUTURE WORK

#### Key Contributions

- Novel visual approach to news aggregation and exploration
- Integrated data pipeline combining multiple news sources
- Interactive interface enhancing news discovery
- Enhanced accessibility for diverse user populations

*Our word cloud dashboard transforms news consumption by distilling complex information into an intuitive visual format, addressing information overload while promoting engagement with current events.*

#### Future Development

Personalization Custom preferences and topic following	Additional Sources More news APIs and social media
Advanced Analytics Sentiment analysis and trend tracking	Mobile Optimization Dedicated app for on-the-go access

### REFERENCES

- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301-2309. <https://doi.org/10.1109/TVCG.2011.185>
- Hanafiah, N., Kevin, A., Sutanto, C., Fiona, N., Arifin, Y., & Hartanto, J. (2017). Text normalization algorithm on Twitter in complaint category. *Procedia Computer Science*, 116, 20-26. <https://doi.org/10.1016/j.procs.2017.10.004>
- Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii International Conference on System Sciences* (pp. 1833-1842). IEEE. <https://doi.org/10.1109/HICSS.2014.231>
- Manning, C. D. (1999). *Foundations of statistical natural language processing*. MIT Press.
- Neuman, W. R., & Guggenheim, L. (2011). The Evolution of Media Effects Theory: A Six-Stage Model of Cumulative Research. *Communication Theory*, 21(2), 169-196. <https://doi.org/10.1111/j.1468-2885.2011.01381.x>
- Tromble, R. (2021). Where have all the data gone? A critical reflection on academic digital research in the post-API age. *Social Media + Society*, 7(1), 205630512198892. <https://doi.org/10.1177/205630512198892>